

Conference Abstract

# High-Throughput Digitisation of Natural History Specimens

Elizabeth Louise L Allan<sup>‡</sup>, Steen Dupont<sup>‡</sup>, Helen Hardy<sup>‡</sup>, Laurence Livermore<sup>‡</sup>, Benjamin Price<sup>‡</sup>, Vincent Stuart Smith<sup>‡</sup>

<sup>‡</sup> Natural History Museum, London, United Kingdom

Corresponding author: Elizabeth Louise L Allan ([louise.allan@nhm.ac.uk](mailto:louise.allan@nhm.ac.uk))

Received: 14 Jun 2019 | Published: 25 Jun 2019

Citation: Allan E, Dupont S, Hardy H, Livermore L, Price B, Smith V (2019) High-Throughput Digitisation of Natural History Specimens. Biodiversity Information Science and Standards 3: e37337. <https://doi.org/10.3897/biss.3.37337>

## Abstract

The Natural History Museum, London (NHM) has embarked on an ambitious Digital Collections Programme to digitise its collections. One aim of the programme has been to improve the workflows and infrastructure needed to support high-throughput digitisation and create comprehensive digital inventories of large scientific collections.

Pilot projects have been carried out for a variety of collection types, from which high-throughput imaging workflows have been developed and refined. These workflows have focused on pinned insect specimens (Blagoderov et al. 2012, Paterson et al. 2016, Blagoderov et al. 2017, Price et al. 2018), microscope slides (whole slide and specimen imaging; Allan et al. 2018, Allan et al. 2019) and herbarium sheets.

The rate and time taken to digitise specimens is influenced by a number of factors that include, among others, the level of preparation and post-processing required, imaging approach, the type of specimens as well as the complexity and condition of the collection. As part of this presentation we will include information on the rate, cost and time to digitise various NHM collections, illustrating how our processes have improved digitisation efficiency and allowed us to maintain quality.

The programme has run a variety of digitisation projects, gathering data about rates of digitisation (preparation, imaging, transcription etc.) and developing improvements.

Collection types such as microscope slides and herbarium sheets lend themselves to higher imaging rates, while other collections such as pinned insects, which require greater amounts of specimen handling to remove labels, tend to have lower imaging rates (Fig. 1). In order to increase efficiency, we have developed approaches that minimise specimen handling. For example, workflows for pinned insects such as the Angled Label Image Capture and Extraction (ALICE) do not require the removal of specimen labels from the pin as the system can capture angled images of the labels, thus increasing the imaging rate three-fold (Fig. 1).

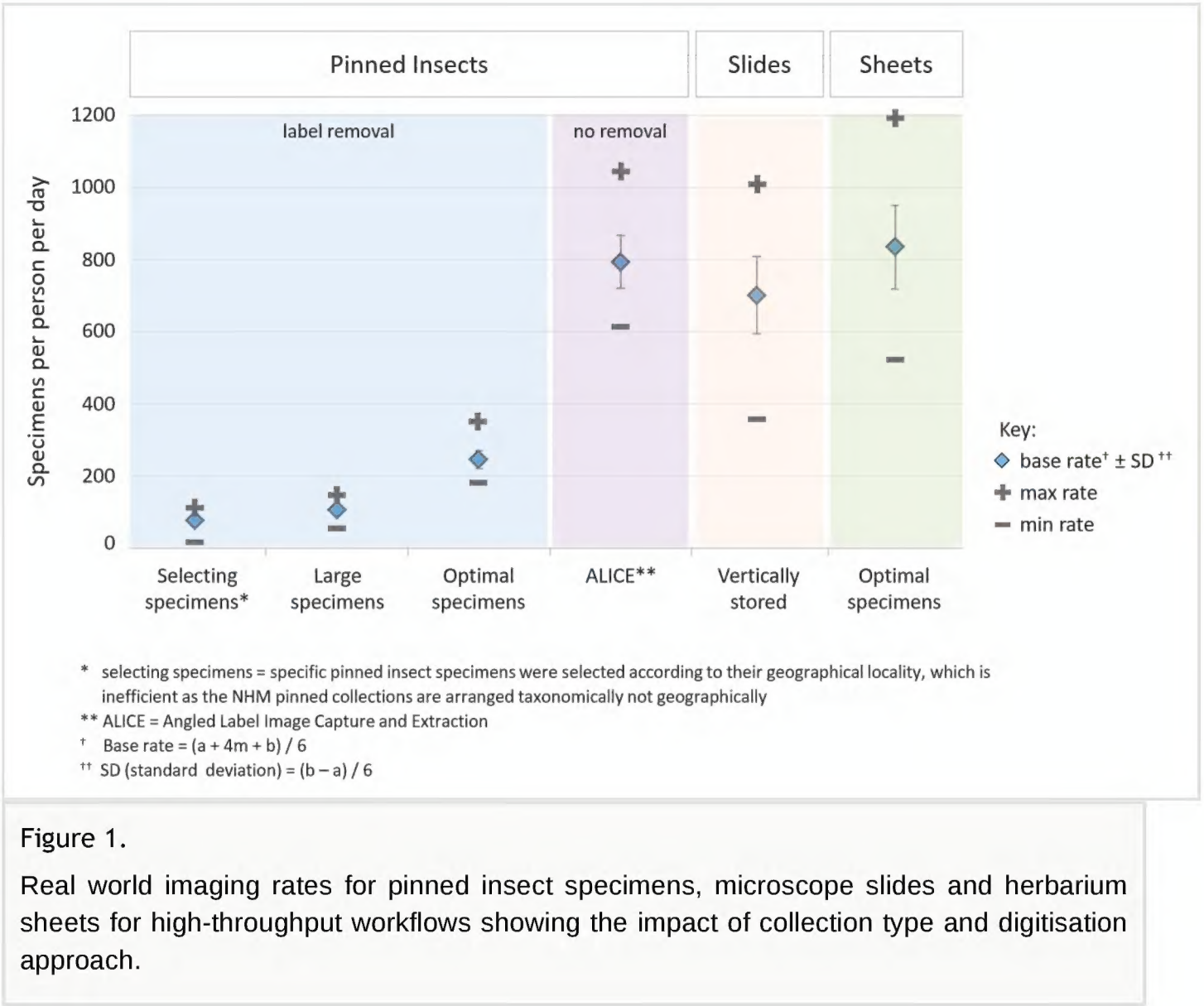


Figure 1. Real world imaging rates for pinned insect specimens, microscope slides and herbarium sheets for high-throughput workflows showing the impact of collection type and digitisation approach.

Another approach taken is to semi-automate mass digitisation using a combination of temporary and permanent Data Matrix barcode labels (Allan et al. 2019). By using multiple barcodes at the imaging stage to encode information associated with each specimen (i.e. unique identifier, location in the collection, taxonomic name, type status etc.; Fig. 2), we can run a series of automated processes, including file renaming, image processing and bulk import into the Museum's collection management system. Through adaptation of our workflows with this new approach we have increased the efficiency of digitisation processes, illustrating how simple activities, like automated file renaming, reduces image post-processing time, minimises human error and can be applied across multiple collection types.



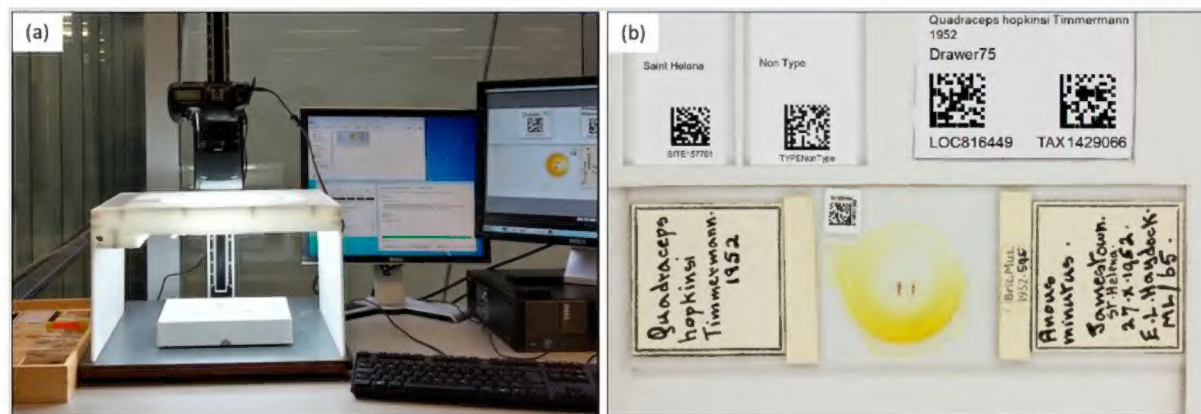


Figure 2.

Imaging specimens using the multiple barcode workflow. (a) Standard imaging setup consisting of a vertically mounted DSLR camera, a custom-built lightbox and an imaging template. (b) Imaging template with a specimen (slide) and various temporary labels with metadata encoded in Data Matrix barcodes, are cropped from the final image (primary key values are used in the CMS to ensure a 1:1 match).

## Keywords

automation, 2D barcodes, natural history collections, mass digitisation, pinned insect specimens, microscope slides, herbarium sheets

## Presenting author

Louise Allan

## Presented at

Biodiversity\_Next 2019

## References

- Allan EL, Price BW, Shchedrina O, Dupont S, Livermore L, Smith V (2018) A low cost approach to specimen level imaging of natural history microscope slides using a DSLR system. OSF Preprint <https://doi.org/10.31219/osf.io/dvmsh>
- Allan EL, Livermore L, Price B, Shchedrina O, Smith V (2019) A novel automated mass digitisation workflow for natural history microscope slides. Biodiversity Data Journal 7: e32342. <https://doi.org/10.3897/bdj.7.e32342>
- Blagoderov V, Kitching I, Livermore L, Simonsen T, Smith V (2012) No specimen left behind: industrial scale digitization of natural history collections. ZooKeys 209: 133-146. <https://doi.org/10.3897/zookeys.209.3178>

- Blagoderov V, Penn M, Sadka M, Hine A, Brooks S, Siebert D, Sleep C, Cafferty S, Cane E, Martin G, Toloni F, Wing P, Chainey J, Duffell L, Huxley R, Ledger S, McLaughlin C, Mazzetta G, Perera J, Crowther R, Douglas L, Durant J, Honey M, Huertas B, Howard T, Carter V, Albuquerque S, Paterson G, Kitching I (2017) iCollections methodology: workflow, results and lessons learned. Biodiversity Data Journal 5: e19893. <https://doi.org/10.3897/bdj.5.e19893>
- Paterson G, Albuquerque S, Blagoderov V, Brooks S, Cafferty S, Cane E, Carter V, Chainey J, Crowther R, Douglas L, Durant J, Duffell L, Hine A, Honey M, Huertas B, Howard T, Huxley R, Kitching I, Ledger S, McLaughlin C, Martin G, Mazzetta G, Penn M, Perera J, Sadka M, Scialabba E, Self A, Siebert D, Sleep C, Toloni F, Wing P (2016) iCollections – Digitising the British and Irish butterflies in the Natural History Museum, London. Biodiversity Data Journal 4: e9559. <https://doi.org/10.3897/bdj.4.e9559>
- Price BW, Dupont S, Allan EL, Blagoderov V, Butcher AJ, Durrant J, Holtzhausen P, Kokkini P, Livermore L, Hardy H, Smith V (2018) ALICE: Angled Label Image Capture and Extraction for high throughput insect specimen digitisation. OSF Preprint <https://doi.org/10.31219/osf.io/s2p73>